

DOCUMENT RESUME

ED 157 943

95

TH 007 364

AUTHOR Lipe, Dewey; Haveman, Jacqueline E.
TITLE Toward More Natural Field Testing of Classroom Materials.
INSTITUTION American Institutes for Research in the Behavioral Sciences, Palo Alto, Calif.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO AIR-47900-6-77-SR
PUB DATE Jun 77
CONTRACT NIE-C-0142
NOTE 37p.; Parts may be marginally legible due to type size

EDRS PRICE MF-\$0.83 HC-\$2.06 Plus Postage.
DESCRIPTORS Career Education; Curriculum Development; *Curriculum Evaluation; Data Analysis; Data Collection; Elementary Secondary Education; *Evaluation Methods; *Field Studies; Formative Evaluation; *Instructional Materials; *Learning Modules; Program Evaluation; Questionnaires; Research Design; Research Methodology; *Research Problems; Research Utilization

ABSTRACT

Issues associated with the evaluation of 61 educational curriculum units, which was conducted by the American Institutes for Research (AIR) are described. The units were adapted to the AIR Career Education Curriculum framework by creating 133 instructional units within which students could develop career skills in an orderly sequence. The evaluation process involved initial revision, field testing, and final revision of the units. During the initial revision, the needs of teachers, school administrators, and publishers were seen to include reduction in the length of units, inclusion of alternative learning activities, reduction in cost, replacement of copyrighted passages, and improvement in the appearance of materials. Obtaining information without arousing resentment or reluctance was described as the most difficult aspect of a naturalistic field test. As a result, teacher orientation workshops were conducted to introduce the curriculum units. School district responses were described as passive resistant, personally committed, aggressive resistant or group committed. Methods for obtaining unambiguous answers to test questions without disrupting the natural setting are described, and sample module and activity evaluation instruments are appended. Results included 454 teacher field test packets and 11,700 student evaluation forms. Computer-assisted handscoring of results is explained, as well as the development of total data sets for each module. (Author/JAC)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

**TOWARD MORE NATURAL
FIELD TESTING OF
CLASSROOM MATERIALS**

TN007 364



AMERICAN INSTITUTES
FOR RESEARCH

This work performed pursuant to Contract No. NIE-C-0142 with the
National Institute of Education.

AIR

AIR-47900-6/77-SR

TOWARD MORE NATURAL FIELD TESTING
OF CLASSROOM MATERIALS

Dewey Lipe

Jacqueline E. Haveman

June 1977

CONTENTS

CHAPTER 1

OVERVIEW OF THE ISSUES	1
----------------------------------	---

CHAPTER 2

WHO ARE THE CONSUMERS, WHAT ARE THEIR NEEDS, AND HOW CAN THEIR NEEDS BE MET?	5
---	---

CHAPTER 3

WHAT DEGREE OF INFRINGEMENT UPON SCHOOL ACTIVITIES IS FEASIBLE, NECESSARY, OR WISE?	15
--	----

CHAPTER 4

HOW CAN A FIELD TEST BE DESIGNED THAT WILL PROVIDE UNAMBIGUOUS ANSWERS TO QUESTIONS WITHOUT DISRUPTING THE NATURAL SETTING?	21
---	----

CHAPTER 5

HOW CAN THE INFORMATION ACQUIRED IN THE FIELD TEST BE MOST EFFECTIVELY ORGANIZED, UNDERSTOOD, AND USED?	29
--	----

APPENDICES

Appendix A: Module Evaluation Form	35
Appendix B: Activity Evaluation Form	37

CHAPTER 1

OVERVIEW OF THE ISSUES

In July 1974, AIR received a shipment of 61 career education curriculum units from the Center for Vocational Education (CVE) at Ohio State University. These units had been developed by teachers from across the country. Their purpose was to illustrate the variety of instructional materials that could be developed, reflecting the school-based curriculum model for career education. CVE had coordinated the development of 133 prototype units. AIR's task was to field test and revise 61 of them.

The purpose of this monograph is to present some lessons that AIR staff learned from this field test effort. It is not the purpose to describe the field test and to present the results. This is presented in a separate document--a final report in four parts.¹ This monograph is the product of staff reflections upon what was done, why it was done, how it worked, and how it might have been done better.

The field test was national in scope and involved 525 classrooms over the course of one school year. These classrooms represented 33 school districts in 16 states. In this respect, it was a large field test effort. Yet it was conducted in such a way as to constitute 130 separate, independent studies. In this respect, then, each study was a modest field test effort. Thus, the lessons learned may apply to national curriculum evaluation studies and to much smaller, local field test designs.

This monograph is organized around four issues that emerged from staff reflections. One issue, perhaps the most fundamental, is the identification of who the curriculum consumers are and what they need in a career education curriculum. Initially, the consumers were thought of as students. The field test objective was to determine whether important career education objectives

¹ Field test report on 130 career education curriculum modules. Palo Alto, California: American Institute for Research, 1976 (AIR-47900-12/76-FR).

were achieved by students via these curriculum materials. Very early in the project, staff was forced to broaden the concept of curriculum consumer to include all of the dissemination audiences interposed between curriculum research and development and student use. The materials have to win the support of each intermediate audience before they will ever reach students. Therefore, the acceptability of the materials to each audience arose as an important question to be answered in the field test.

The second issue concerns how much staff would want schools to change their normal operating and instructional procedures in order to accommodate the curriculum materials evaluation design. Without compromising either the purity of the evaluation design or the school's procedures, normally there could be no field test at all. For example, random assignment of students to treatment and control classes would typically require disrupting the scheduling procedures and changing some students' course of studies. Since such severe disruption is usually unacceptable, the design must be compromised to deal with intact classes.

Staff questioned whether purity of design was something that was desired even if it could be obtained. Shouldn't the concern instead be with how well the program works under natural school conditions, unfettered by design constraints? Isn't it possible that researchers' intrusive influence confounds field test results as much as the confusion resulting from compromises in the evaluation design?

Staff chose to conduct the field test under natural school conditions rather than to try to shape nature into conformity with the design. In a naturalistic field study, research staff must not be perceived as infringing upon the domain of responsibility of school staff. While this could not be achieved in the ideal, it proved to be a goal worth pursuing, the benefits of which make naturalistic curriculum evaluation an issue worth considering here.

Given a naturalistic approach to field testing, the third issue followed from resolution of the second issue. How could data be obtained to answer the research questions unequivocally? This issue seemed to return to the

question about research design. After all, answering questions unequivocally is the objective of research design. Thus, the third issue concerns how to design the field test to fit the natural conditions within schools. An unconventional design that worked very nicely will be described.

The fourth issue concerns processing and analyzing thousands of pieces of information from the field test and using them to guide final revision. Instead of designing a complicated series of automated computer processes, much of the data processing was done by hand, using the computer to perform only the scoring/item analysis procedures and analysis of covariance. Although there are persuasive arguments for both approaches, hand processing was selected because staff wanted to see and synthesize raw data firsthand. How this specifically enhanced the final revision task will be discussed.

CHAPTER 2

WHO ARE THE CONSUMERS, WHAT ARE THEIR NEEDS, AND HOW CAN THEIR NEEDS BE MET?

Within two weeks of the beginning of the project, the first major contingency arose. It led AIR, in conference with NIE sponsors, to introduce the first and most comprehensive change in the task to be done. Staff learned that no publisher was willing to take on the publication of a group of units similar to the ones received that had already been field tested and revised. Whereas staff had taken for granted that the field testing and revision of the units would ultimately result in commercial publication, this assumption became clouded with serious doubt. This led staff to question whether the units were ready for a major field test effort.

Commercial publication was desired because major commercial publishing companies advertise, promote, and actively solicit the sale of their products. Without this orchestrated promotional backing, educational R&D products don't get widely used. Thus, the ultimate success of the dissemination of this career education curriculum came into question.

From AIR's point of view as the new group taking on the continued research and development of 61 units, it was necessary to look again at who, indeed, were the consumers of the curriculum materials. Students may be the target consumer, but before reaching the students the product must be accepted by a variety of intervening audiences. The materials must be accepted by teachers who will teach them effectively. In order to reach teachers, they must be chosen for purchase by school administrators who make such purchasing decisions. Prior to that, they must be brought to the attention of those administrators, presumably by a field representative of a commercial publisher.

In order to assess the acceptability of the units to these three intermediate audiences, AIR interposed an acceptability study as a necessary first step in the project. Representative units were sent to a random sample of school administrators across the country. Also three workshops were held for

experienced career education teachers and curriculum specialists who examined and critiqued all of the units. And 31 publishers of career education materials were contacted. With 15 of those publishers who expressed strong interest in the materials, staff personally discussed possible publication issues and options.

As might be expected, some needs conflicted across these three groups. Publishers needed to show a profit, and some of them were looking for student consumable materials to provide continuing income from the materials. Administrators wanted to cut expenses. Teachers wanted to use less paper by eliminating much of the white space in the printed materials and limiting the amount of student consumable materials.

Nevertheless, many of the needs of the three groups proved compatible, even complementary. The following 12 points represent those compatible and/or complementary needs.

- 1) **Reduce the Length of Units**
Curriculum specialists and teachers agreed that the units were too long--i.e., more class time was required for most units than teachers generally wanted to devote to the career education topic being taught.
- 2) **Partition Units into 120-180 Mini-units**
Teachers indicated that the units would be more practical (more easily implemented) if they were partitioned into homogeneous subunits. Then a subunit could be taught in its entirety within a relatively short time interval, to be followed at a much later time, if so desired, by another subunit. This would allow greater flexibility in using the material over a long period of time by providing for closure at interim points in time.
- 3) **Provide for Alternative Learning Activities**
A major concern of all sources surveyed--curriculum specialists, teachers, and publishers--was that particular learning activities were over-emphasized at particular grade levels. For example, teachers indicated that too much emphasis was put upon large-group activities and class discussion in grades K-3; less writing with more roleplaying and community involvement were needed in grades 4-6; less reading with more graphics in student materials and more "manipulating" and "doing" were needed in grades 7-9. Thus, alternative learning activities needed to be supplied in order to provide for individual differences and to allow flexibility and adaptability in the use of the materials.
- 4) **Reduce the Cost of Each Unit**
Curriculum experts and teachers insisted that the cost of each unit must be low. Learning activities that depended upon expensive resources, including copyrighted resource materials, had to be replaced by less expensive means.

- 5) **Replace Copyrighted Passages**
Of primary concern to publishers was the fact that there were 24 units which had been tentatively identified as having potential copyright problems. These passages in the Teacher's Guide or student materials needed to be replaced.
- 6) **Improve the Appearance of Student Materials**
Curriculum specialists, teachers, and publishers were in consensus with regard to the need to make the student materials more appealing visually.
- 7) **Expand the Occupational Base of the Units**
Teachers and publishers expressed concern that the occupational clusters were not covered broadly enough.
- 8) **More Carefully Adjust the Conceptual and Reading Levels to the Targeted Grades**
A concern of all sources surveyed was that the language be simplified, reading levels generally reduced, and concepts in some cases be covered more thoroughly.
- 9) **Carry Out Miscellaneous Minor Editing**
Various nonsubstantive editorial changes were required, such as targeting a unit for a grade range rather than a specific grade level and making the Teacher's Guide more concise.
- 10) **Package the Units in Marketable Fashion**
Curriculum specialists, teachers, and publishers were all dissatisfied with the packaging of the units; however, there was not a consensus on the most marketable format. Teachers indicated that they liked the idea of having masters of student materials. And, they wanted to be able to select the lessons that they liked best from a unit.
- 11) **Provide More Objective Evaluation Procedures**
Curriculum specialists and teachers wanted evaluation to consist of more than a teacher certification that students had done a particular task adequately.
- 12) **Link Units into a Coherent Program**
Publishers, in particular, were concerned that the objectives be consistent across grade levels and that they show a progression of skills and attitudes. They did not need to form a comprehensive curriculum, but they did need to lead toward essential career development goals.

In summary, there were gaps to be filled in and emphases to be shifted. A major task was to reconfigure and augment the units to form a coherent program comprised of smaller, independent units. Much of the material in the units was imaginative, well thought out, and usable; however, substantial revision was clearly indicated as a task to precede the field test. As a consequence of these results, the sequence of field testing and then revising the units, as originally planned, was resequenced and expanded to include revision, field testing, and final revision.

The first eleven consumer needs called for careful and creative revision of the units; however, they did not pose insurmountable problems. The twelfth point did present a dilemma. The dilemma was how to make a coherent program out of units that were not intended to comprise such a program. The original 133 units were designed to illustrate major elements of a career education curriculum. They were not intended to be comprehensive. Furthermore, AIR was to field test and revise fewer than half of the original set. This gave even less potential for ferreting out a coherent program from the 61 units AIR had.

As point number twelve was examined in relation to the other eleven recommendations, staff saw that points 1, 2, 3, 7, 10, and 11 could be considered aspects of the same problem. The length of each unit (number 1) could be reduced by partitioning it into separate subunits (number 2). Then additional subunits could be developed, if necessary, to give adequate coverage of the occupational clusters (number 7). Learning activities could be revised, deleted, and added to meet the several concerns in number 3 and also to make a functional and marketable product (number 10). In the process of doing all this, staff needed a structure to assure that this backing and filling would result in a coherent program for grades K-12.

The structure chosen was one developed and field tested at AIR and found highly functional for infusing career education into the typical existing school curriculum. This AIR career education curriculum framework is organized into levels of increasing detail so that the level of specificity can be picked that best fits the amount of material desired in the completed program. The AIR Career Education Curriculum Topic Chart in Figure 1 shows the first four levels and how each category in a given level is further subdivided at the next level of detail. There are two levels of greater detail not shown in Figure 1. Each topic is further divided into concepts, and each concept is comprised of a set of instructional objectives. There are 178 concepts and 1,981 objectives.

In order to assess whether the 61 units could be represented by the AIR Career Education Curriculum framework, the following steps were performed:

- 1) Partition the 61 units into their 716 component lessons. Lessons that required more than an hour of class time were split into sessions so that the adjusted number of lessons came to 859.

- 2) Analyze the content of each lesson to determine whether it can be represented in one of the 31 topic categories of the AIR Career Education Curriculum Topic Chart.
- 3) Place each lesson in its appropriate grade range: K-1, 2-3, 4-6, 7-9, or 10-12.
[The result of step 2 and step 3 was a matrix of 155 cells. Thirty-one curriculum topics comprised one dimension, and the five grade ranges comprised the other.]
- 4) Identify which concepts from the AIR Career Education Curriculum were covered by the lessons in each cell. Add and/or subtract concepts and realign curriculum areas in order to form a coherent program.
[A coherent program was defined as one leading toward essential, pre-entry level career goals.]
- 5) Group the concepts within each cell so that each group can be developed into an independent instructional module.

The result of these five steps was a Career Education Program in three strands and eleven areas as shown in Figure 2. It provided the framework for a program consisting of 133 instructional modules within which students could develop essential career skills in an orderly sequence.

Having developed the programmatic structure for these curriculum materials, an extensive revision task remained to be accomplished. That task entailed revising and augmenting lessons associated with each group of concepts and adding new lessons as necessary to complete each module.

What lessons, then, did staff learn with regard to this first issue? Staff learned that it pays to ask certain questions even though the answer at first appears obvious. Issue number one would not have become an issue had not the question been asked, "Who are the curriculum customers?" And the question would not have been asked if events early in the project had not revealed a lack of publisher commitment to the units. Many publishers were interested, but not one of them was ready to put up the money to back their expressed interest.

Even at this point, however, the issue might have been ignored, simply assuming that the field test and revision effort would make everything right. AIR had a good plan of action already accepted under contract with NIE, and staff might have argued for keeping that original plan. By asking something

Figure 1

THE AIR CAREER EDUCATION

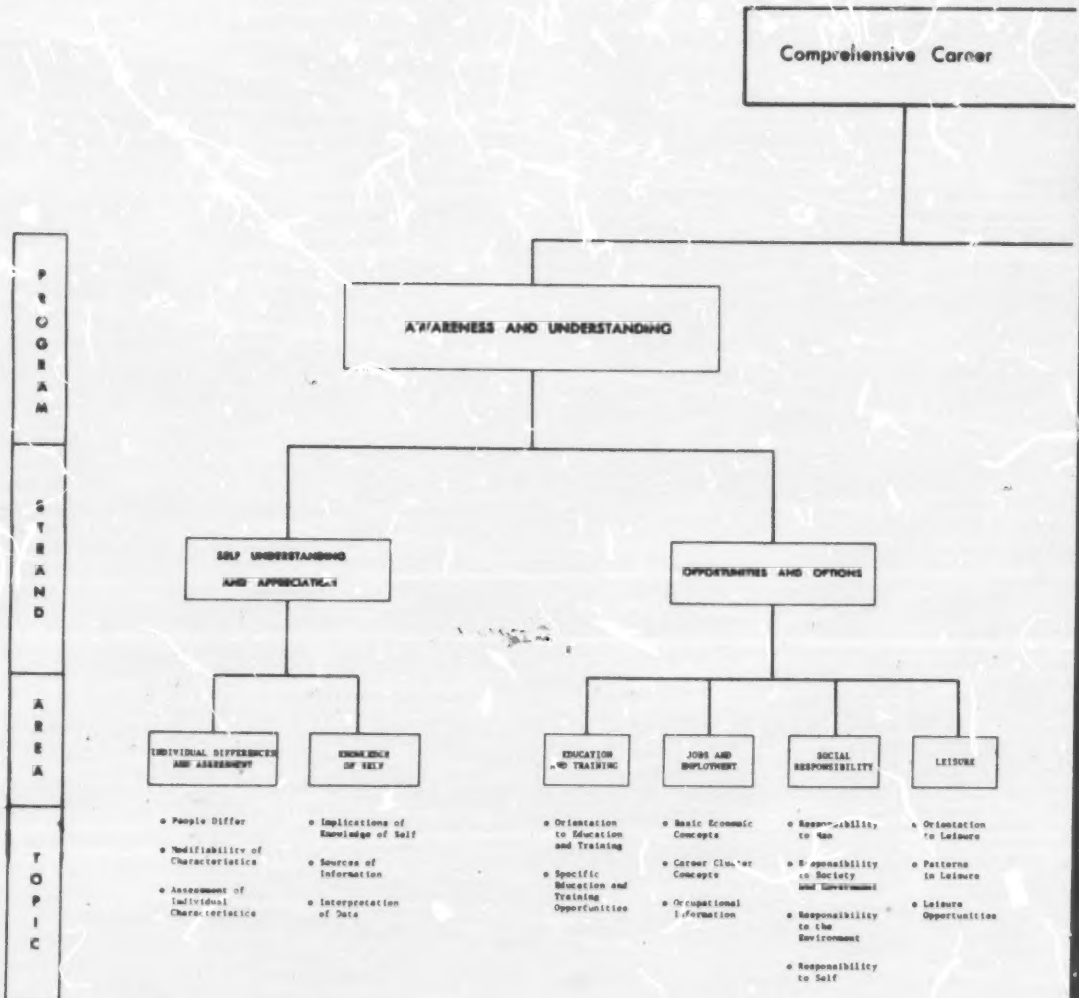


Figure 1 (continued)

CURRICULUM TOPIC CHART

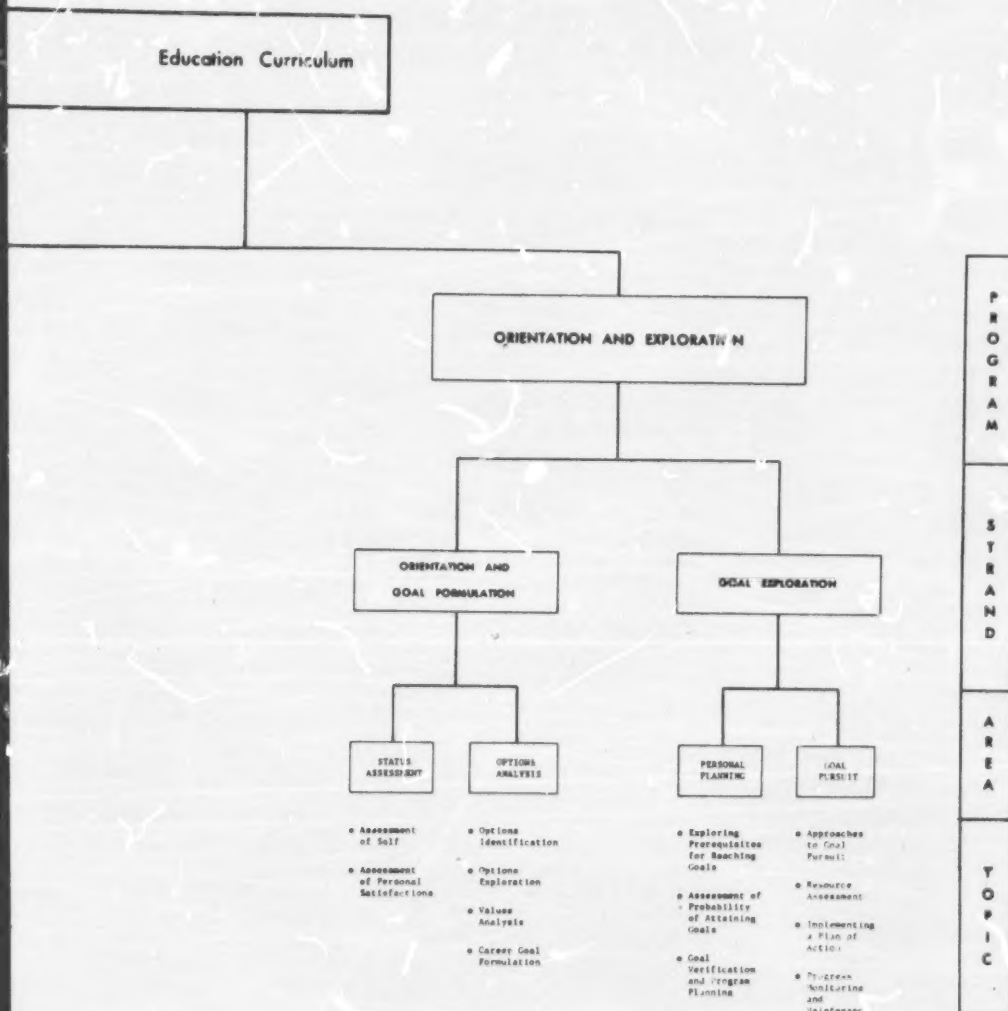
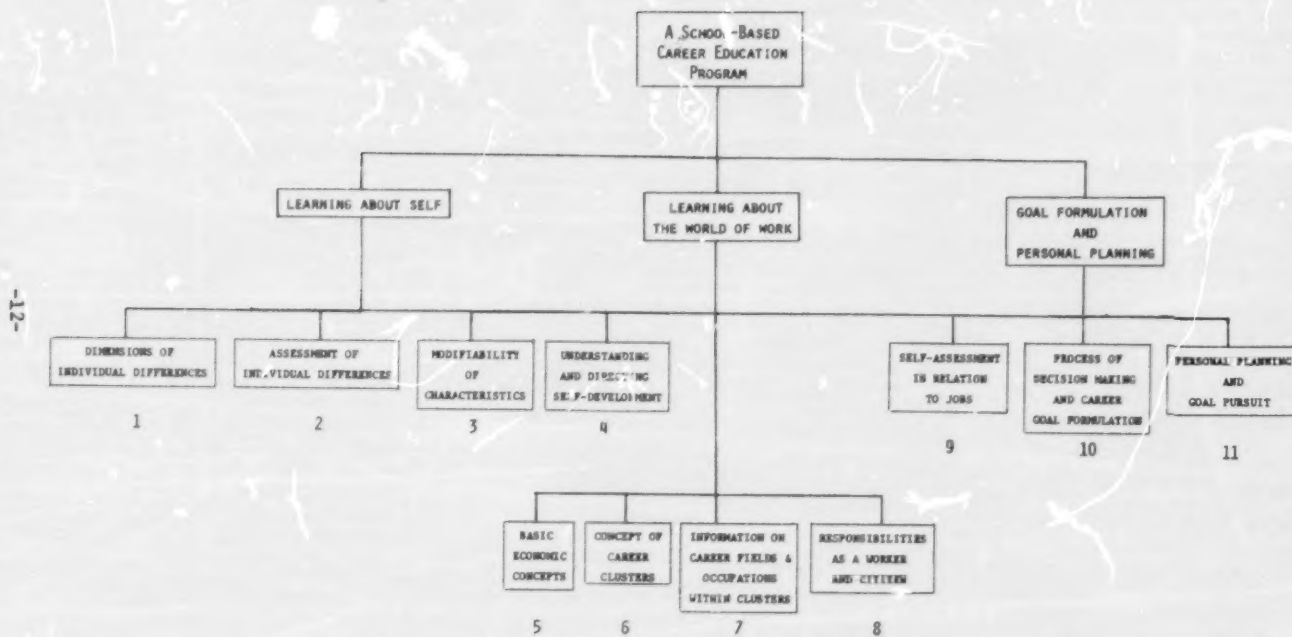


FIGURE 2



as basic as who the curriculum consumer is, AIR ran the risk of having to change the basic strategy for the project. This meant that a period of uncertainty would exist until a new strategy could be formulated.

Staff learned that a mental set of creative flexibility along with sound decision-making strategy is appropriate even for a large project such as this. Staff had to loosen the established concept concerning who the consumers were and then had to be willing to add an information-seeking task in order to find out what the new set of consumers needed. Staff was then required to suspend further development of strategy until the results of the information survey were received and synthesized. After the results were in, the need became apparent for a drastic change in the original plan.

The resolution of issue number one prepared the way for a combined revision/development task that hopefully would result in a more relevant, usable, effective career education program than would otherwise have been possible.

CHAPTER 3

WHAT DEGREE OF INFRINGEMENT UPON SCHOOL ACTIVITIES IS FEASIBLE, NECESSARY, OR WISE?

Ideally, staff might have wanted to field test the revised career education modules in a stratified random sample of U.S. public and private schools where they would be taught by randomly selected teachers to randomly sampled students. Then the results could be generalized to the U.S. population of schools, teachers, and students. On the other hand, the constraints of "reality" pointed to several reasons why this might not be feasible to pursue.

In the first place, it was not likely to be achieved; no one had done it in the past. Schools are not laboratories where tight controls can be exercised. In the second place, even if it were achievable, would it be desirable for project purposes? If the program were to be distributed by a commercial publisher, the target schools would probably be those whose curriculum administrators had some initial interest in career education. As supplementary materials, the modules would likely be taught on a voluntary basis rather than prescribed by the school board. Or, if prescribed as a program, the teachers would likely select only those modules that they thought interesting and appropriate to their students. In sum, schools would be self-selected and teachers would have some degree of choice in their use of the materials.

A field test that distorted these and other elements that constitute a "natural" environment for the materials would not provide the necessary information. Staff needed to know whether the materials would be used and, if used, whether they would be used effectively to promote desired student outcomes. The following elements were identified that seemed to represent the minimum essential characteristics of a "naturalistic" test of the materials:

- The school district curriculum administrator and the school principal would ask to have this career education program tested in their school.
- Who teaches the materials would be determined by the curriculum administrator and the school principal. AIR asked only that at least two teachers be included from any one of the five grade ranges.

- Participating teachers would choose which modules to teach and which lessons within modules they would use.
- Teachers would change, adjust, or delete parts of lessons as they wished; however, they were asked to note and explain any changes they made.

The first step was to inform school administrators about the field test. A general invitation was issued to school districts through the AIR Career Education News, through announcements at professional meetings, and directly to school administrators with whom staff had worked in earlier projects. School districts that had been randomly selected for the acceptability survey were also invited. Ninety percent of the latter group responded that they were interested in participating in the field test of the materials. The level of response was indicative of the interest in career education among school administrators in 1974.

There were, of course some logistical and other practical considerations to be resolved in the process of firming up the agreement to include a school in the field test. More schools were included than were needed to adequately field test the modules. This was because some school administrators expressed so strong a desire to be included that staff decided to increase the scope of the field test, within budgetary limits, to serve that need.

After commitment to participate had been established, the next step was to arrange for a teacher orientation workshop. This was typically a two-hour, in-service session in which an AIR staff member described the program, reviewed the evaluation requirements, and then displayed modules from which interested teachers selected the ones they would like to teach. The orientation session, whether attended by three or four teachers or as many as 50 teachers, was informal. Essential information was presented concisely, and the remainder of the time was spent answering individual teachers' questions as they reviewed the modules.

Each participating school district was unique in its involvement with the materials and with AIR staff. This uniqueness could be identified on dozens of attitudinal and behavioral dimensions. One large school district had arranged for television news coverage of the district's field test participation, including

an interview with the AIR staff member. In another large district, only half a dozen teachers were present, and they were angry and hostile and, in the end, chose not to participate. In other districts, the response was overwhelmingly enthusiastic.

The tone of schools' responses to the teacher orientation workshops is described here because it pertains directly to the issue of infringement caused by the field test. From AIR's point of view, infringement must be defined as a matter of perception on the part of the school staff. It cannot be described in terms of new demands upon teachers nor in terms of procedural changes in teachers' routine because these situations can occur without any intrusion from an outsider wanting to field test new materials. Furthermore, these situations may be perceived as "normal" intrusive events quite apart from any involvement of an outside agent.

What perceptions, then, would be interpreted as constituting infringement upon school activities? For AIR purposes, the field test constitutes an infringement to the degree that it is perceived as intrusive and to the degree that the intrusion is perceived as being caused by an outside agent. The teacher orientation workshops constituted the only face-to-face encounter between AIR and school staff. It was the one opportunity for our staff to judge firsthand the teachers' perceptions of the field test effort. On the basis of those informal judgments, AIR designated four qualitatively different perceptions among the school districts where teacher orientation workshops were conducted.

- 1) All or nearly all teachers in the school were present. They sat quietly and responded minimally. They appeared to have come as a result of borderline coercion on the part of their administrators.

This category might be labeled as passive resistant.

- 2) Few teachers were present, and they were all highly committed to and enthusiastic about participating. They felt that career education was very important. They wanted the program to prove highly effective so that other teachers would become interested enough to introduce career education in their classes.

This category might be labeled as personally committed.

- 3) Few teachers were present, and they were openly hostile. They expressed anger over being told and resentment toward the AIR staff member for presuming to tell them what to do.

This category might be labeled as aggressive resistant.

- 4) Nearly all the teachers from the school were present, and they exuded enthusiasm. Career education was a high priority commitment of the school that apparently was shared by all the teachers present.

This category might be labeled as group committed.

The description of each of these categories, especially that of number three, is exaggerated in order to accentuate the "character" of that category. Obviously, category three constitutes a high infringement by an outside agent. Category one is less drastic infringement; if AIR had not happened to come along, some other means would have been devised by the administration to introduce career education into their curriculum. Neither category two nor category four constitutes an infringement because the participants did not perceive the field test as intrusive.

The question, then, is whether the curriculum to be tested can be introduced naturalistically and whether essential data will be willingly provided. The first part of this question will be considered here under issue two, and the second part will be considered under issue three.

In reflecting upon AIR involvement with schools, staff asked whether, in retrospect, it had succeeded in introducing this career education program naturalistically into the instructional program of those schools. What would be done differently next time? "Naturalistic" was defined as being a matter of perception that the field test was not an intrusion by an external agent. Experimentation with innovative techniques and materials in many schools is so common as to constitute an essential feature of those natural settings. Thus, newness of the curriculum cannot categorically default the field test as unnaturalistic. One feature appears to be all important in allowing the field test to be unobtrusive. That feature is "choice."

School administrators were invited to participate; they were not asked to participate. The difference is in the fine line of no implied sales pitch, no implied sense of urgency, nor any undertone of implication that the school would be doing AIR a favor by participating. The invitation was worded so that the choice of whether to participate was unencumbered by obligation or pressure.

The next choice situation was the one in which the school administrator organized the teachers to participate in the teacher orientation workshop. Because AIR did not want to interfere in any way, staff did not give any instructions to administrators about how they should obtain teachers, nor did staff urge them to have lots of teachers present. The omission of any guidelines in this matter, however, left the administrator the option of telling teachers that AIR wanted to do a field test. This would place the focus on AIR and the field test rather than on the new curriculum materials that teachers might want to use.

This aspect would be handled differently next time. The administrator would be asked to introduce the career education program to teachers in the same way he or she would introduce any new curriculum materials. The fact that AIR was conducting a field test would not be hidden; it would only be relegated to a position of secondary importance to the teachers. After the teacher orientation workshop had been held, AIR staff would ask that the administrator briefly describe how the teachers had been approached. This would be done as a check on the assumptions about the naturalness of the field test.

The teachers' and students' involvement was naturalistic, in AIR's view. Staff suspect that it would be naturalistic in any curriculum field test classroom, regardless of the written directions to teachers. If teachers believe they can improve instruction by deviating from their field test guidelines, they will do so. Permission was given specifically for the teachers to do what staff suspected many of them would do anyhow. They were asked to change the module in any way they thought would improve it with regard to their students and to jot down how they changed it along with a brief explanation.

Teachers were given a month to complete their module, which was about four times the minimum time they would need. A special telephone line was installed at AIR for use by the field test staff, and teachers were asked to telephone collect if they had any questions or concerns.

The most difficult aspect of a naturalistic curriculum field test is to obtain evaluation information without arousing resentment or inducing a feeling of reluctance. This aspect is discussed next as a separate issue.

CHAPTER 4

HOW CAN A FIELD TEST BE DESIGNED THAT WILL PROVIDE UNAMBIGUOUS ANSWERS TO QUESTIONS WITHOUT DISRUPTING THE NATURAL SETTING?

AIR's goal was for teachers to participate in the field test and to do so without perceiving AIR as a source of infringement upon their instructional activities. Low-key involvement with schools and staff efforts to maximize the number of opportunities for teachers to choose alternative courses of action seemed to achieve this goal. Staff did provide the minimum support service that teachers typically expect with regard to any new instructional program. That service was to present an orientation to the program and to be available for consultation concerning any problems or questions that might arise as teachers use the program.

The teachers' responses throughout the field test indicated that they understood the program, that the logistics were adequate, and that the program was being implemented naturalistically, by AIR's definition. The school staff was satisfied up to this point; however, the question remained as to whether AIR's evaluation requirements would be met. Evaluation procedures needed to be developed that would tell whether the program was working; if not, why not; and how the program could be made better. More specifically, four questions had to be answered with regard to each module:

- 1) Was the module infusible into the regular school program with a minimum of teacher training and teacher effort?
- 2) Were the learning activities manageable by students and teachers?
- 3) Were the learning activities interesting to students?
- 4) Did students achieve the desired learning outcomes?

If a module was not doing one or more of these things, information was needed that indicated why, and then corrective changes needed to be made in the module. The latter required that ways of improving the modules be pinpointed during final revision, after the field test.

Obviously, one open-ended question such as, "How can this module be improved?" would not get the specific information needed. Nor would teachers and students take the time to answer many pages of detailed questions. The respondents' thinking needed to be directed and focused and lucid responses facilitated. This needed to be done unoffensively and without inhibiting the respondents' natural style.

While directing the respondents' attention to the evaluation parameters of concern, staff also wanted them to record their insights when those insights occurred and to avoid distractors such as concern about the beauty of their penmanship or about the flawlessness of their sentence structure. In sum, staff wanted to obtain the necessary evaluation data and to have the data flow naturally from the field test activities.

Of the four evaluation questions stated earlier, the first two required information from teachers. The first question was an assessment of the module as a whole, and the second concerned each learning activity (lesson) separately within the module.

To the first question of whether the module was infusible, six areas of concern were identified with several specific items in each area. This could easily have escalated into a 50-page questionnaire; therefore, a limit of one page was established, front and back if necessary. A second self-imposed limit was that a teacher be able to complete the form in five minutes or less. These two limitations ruled out almost every response format except concise stimulus phrases and checked (✓) responses.

Each of the six areas of concern was phrased as a brief question that was followed by up to six key words or phrases, each followed by a five-point scale with three anchoring phrases. One of the questions with six scales is shown below for illustration. (The complete Module Evaluation Form is in Appendix A.) Space for comments was provided under each question where respondents could add any explanation to help in upgrading the module.

C. What is the quality of the materials?

layout and pictorial matter:	()	()	()	()	()
	stimulating		satisfactory		uninviting
accuracy of information:	()	()	()	()	()
	accurate		minor misrepresentation		major misrepresentation and errors
bias:	()	()	()	()	()
	equally attractive to all subgroups		some bias		slanted toward or against one sex or group
comprehensiveness:	()	()	()	()	()
	adequate coverage of topic		neglect of some aspects		large inadequacies
clarity:	()	()	()	()	()
	clear		ambiguous		confusing
evaluation materials:	()	()	()	()	()
	reflect module content		somewhat reflect content		fail to reflect content

COMMENTS

Staff learned from the initial tryout of this format that teachers would answer the questions and occasionally comment; however, many of their comments were general and not very helpful. For one thing, staff needed to know what specific parts of the module required change.

The solution was to send two copies of the Teacher Guide and ask the teacher to write her or his comments in the margin of one of them. That Guide could then be returned to AIR along with the other evaluation forms, and the teacher could keep the clean copy of the Guide. This procedure did produce the specificity desired.

The second of the four evaluation questions stated earlier was also to be answered by teachers, and it concerned whether each of the learning activities (lessons) was manageable. Most of the modules had from three to six learning activities, and a separate evaluation of each one was needed. Clearly, the Activity Evaluation Form had to be contained within even more restricted limits than those imposed upon the Module Evaluation Form. The Activity Evaluation was restricted to one side of one page and the completion time to about three minutes.

One indicator of manageability was the amount of teacher preparation time and amount of student preparation time on each activity. This was to be estimated

retrospectively after completing each activity. If the teacher omitted the activity, he or she was asked to check why. If the activity was modified, he or she was asked to check how. Three scales concerned the activity's usefulness, interest, and acceptability to students as judged by the teacher. Again, space was provided for comments. (The Activity Evaluation Form is in Appendix B.)

The third of the four evaluation questions stated earlier concerned how interesting the module was to students. Teachers indicated their assessment of student interest on each Activity Evaluation Form. In addition, students were asked to report their own level of interest in the module as a whole.

Staff wanted spontaneous and candid responses from students. To achieve such responses, stimulus questions had to be simple, short, and unambiguous. They also had to represent phrasing that students might use themselves. In constructing the questions to assess students' reactions to a module, staff asked a group of local elementary school children to comment on recent classroom activities. From their responses, a pool of twenty statements was constructed, approximately half positive and half negative.

The twenty statements were typed on cards and given to a panel of judges consisting of students in grades 2-6. The judges sorted the statements into five piles, under categories ranging from very good to very bad. Six statements were selected as being most representative of the range of opinion and having the smallest variability (least ambiguity).

Two different response formats were tried out with the initial draft of the form. A five-point Likert format and a simplified three-point response scale were field tested on two small samples of students. Discrimination indices were calculated for each item in both formats, and the coefficient alpha reliability determined. The reliabilities were both above .70, and the discrimination indices were similar; however, the teachers expressed a preference for the simplified three-point response format. The final version of the scale follows:

- | | | | |
|------------------------------------|-----|-------|----|
| 1. This was really interesting. | Yes | Maybe | No |
| 2. It was kind of dumb | Yes | Maybe | No |
| 3. I want to do it again. | Yes | Maybe | No |
| 4. I hated doing this. | Yes | Maybe | No |
| 5. I sure learned a lot from this. | Yes | Maybe | No |
| 6. It was too hard for me. | Yes | Maybe | No |

How could it be made better?

The fourth and last of the evaluation questions stated earlier concerned whether students achieved the learning outcomes intended. Since evaluation is fundamental to instruction, there was no problem associated with the naturalistic assessment of student achievement. Each module had an assessment instrument to measure student attainment of that module's instructional objectives. The module test typically had four items keyed to each of the module objectives. Since a module usually had three or four objectives, the test was twelve or sixteen items long in most modules.

The goal was to answer the fourth evaluation question unambiguously but within the design constraints of a naturalistic field test. A simple posttest score would not tell how much students knew before they had done any of the module's learning activities. This concern could be answered by administering a pretest. Giving a pretest would not constitute a severe imposition upon students and teacher because the pretest scores could be used diagnostically to shape instruction to class needs. In other words, the pretest, too, could be made inherent to module instruction.

The next level of refinement that was considered for inclusion in the design placed somewhat greater strain on the naturalistic parameters of the field test setting. That refinement was to determine whether students might have achieved a given module's instructional objectives as a result of factors other than the module itself. Perhaps a teacher's enthusiasm concerning career education is the primary contributing factor. Perhaps just participating in this career education program, regardless of which module is taught, could account for outcome scores on a given module. To answer this concern, staff

considered comparing outcomes on a given module with outcomes of a control class. The problem was where to get an adequate control class.

After designing several possible courses of action, staff decided upon one that would enable the addition of this refinement with a minimum of further burden on teachers or students. The chosen solution was to have each class in which a module was taught also serve as a control for another class in the same school and in the same grade range. This provided some assurance that the control teacher had a similar level of commitment to career education as the treatment teacher. It also provided a control class in which students were studying a different module within the same career education program.

By adding the feature of control classes to the design, staff also added three constraints upon the field test setting, as follows:

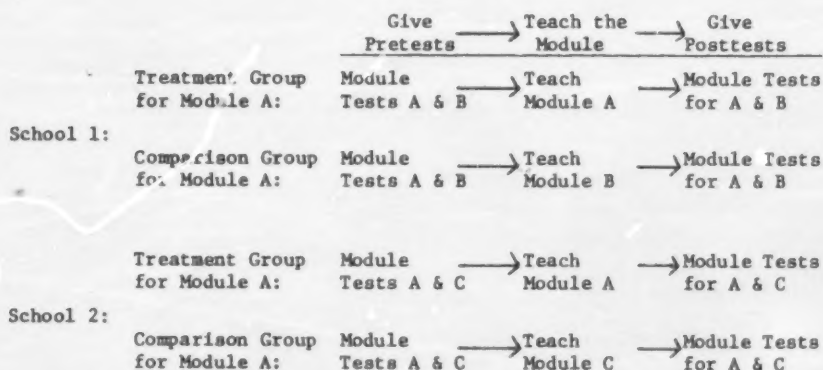
- 1) At least two teachers needed to participate from a given school and from a given grade range.
- 2) Teachers and students needed to be willing to double the length of the pre- and posttest.
- 3) Teachers needed to explain to students that one test was unrelated to their module, that they might find the test interesting anyhow, and that they should see how well they could do on the test even without studying the module.

Even with these constraints, this approach proved to be workable. In each class, the control test was stapled to the test for the module being taught. The items were all multiple choice, and therefore an additional 12- or 16-item test required little more testing time. The content of the control test was career education and, therefore, related to the general area of study as that of the module being taught.

This approach also served the control needs of the field test. Two teachers from a given school were very likely responding to the same kinds of incentives to participate in the field test and therefore were likely to be similarly motivated. Control students were similar to the treatment students in their level of involvement in career education in that the control class was also a treatment class for a different module in the same field test. In short, this design placed a heavy burden upon each module to prove itself

effective. A module was unlikely to sneak by for reasons other than its own merits. The design is shown in paradigm form in Figure 3.

Figure 3
FIELD TEST DESIGN



With these data, staff could see how well students in the four classrooms did on the test of Module A before studying any module. In the two treatment classrooms, staff could see how well students did after receiving instruction in Module A. These results could be examined in relation to those of students in the two comparison classrooms who studied Module B or C and then took the test of Module A. Since the sex of each student was identified on the module tests, staff could also make all of the above comparisons separately by sex. Analysis of covariance was AIR's statistical test in which pretest scores provided the covariate, and main effects included treatment, sex, and school.

In retrospect, staff now asks whether unambiguous answers to questions were achieved without disrupting the natural setting of the materials instruction. The answer is "no, but" AIR came closer than had previously been expected. The instruments were highly functional. There were no indications that the instruments were cumbersome or unclear. AIR was able to include control classes in the design without threatening the naturally occurring instructional process.

CHAPTER 5

HOW CAN THE INFORMATION ACQUIRED IN THE FIELD TEST BE MOST EFFECTIVELY ORGANIZED, UNDERSTOOD, AND USED?

Staff anticipated that the field test, once set in motion, would begin churning out data in staggering quantities for the small field test staff. By design, these data did not pour in all at once. Instead, they flowed in steadily over the period of a school year. Nevertheless, our anticipation of the total volume of data was supported in fact. Out of 525 teachers who received a field test packet following their teacher orientation workshops, 454 returned complete data sets. In addition to the evaluation forms filled out by teachers, data from approximately 11,700 students were also returned. Each student took both a treatment and a control pretest and posttest; thus, over 46,000 module tests were returned in addition to the approximately 11,000 Student Evaluation Forms.

The field test staff was small, and its responsibilities extended far beyond just processing and analyzing the returned data. The teacher orientation sessions were scheduled and conducted by field test staff throughout the school year. After each orientation session, the materials required of each teacher were assembled, packaged, and shipped from AIR. Questions and concerns of teachers were personally answered by the field test staff by phone and followed up, as necessary, by additional mailed material. As the data were returned, the staff stamped code numbers on each test and evaluation form to preserve the anonymity of teachers and students. Then we sent a thank-you note to each teacher. With all these tasks to be done by a staff of five persons, it seemed obvious that either additional staff would have to be obtained or else maximum use of automated computer procedures would be required. As it turned out, neither of these alternatives provided the optimum solution, as explained below.

The above points were practical issues to be resolved but were not the central factors to be considered in deciding how the data should be processed.

Our guidelines for processing and analyzing the data were determined in the first instance by the purpose the data were to serve. The data from the field test were to be used by the revision staff in performing the final revision of the modules. Each module, then, was to become a focus of the data processing and analysis and 130 data packets would result (one packet for each module). Within each packet, the data from each classroom would be presented intact as well as summarized across classrooms. This would preserve the variability of results and enable the revision staff to treat each classroom as a separate replication study of that module.

The field test staff and the revision staff collaborated in designing the contents of the field test packets. The packet was to provide its users with a concise picture of implementation problems and outcome effectiveness of the module. Comments and ratings made by the teachers and students and the module test results were to be presented in such a way that comparisons could be made across classrooms. While reading the field test results on a given module, a revision staff member would first evolve some hypotheses about the weaknesses and strengths of the module and about how it could be improved. Then he or she would go back over the data to see how well and how consistently each hypothesis was supported by the data.

Another data source that surfaced serendipitously was the field test staff itself. This group came to know the school sites and teachers rather well. They also knew the contents of the modules in considerable detail as a consequence of describing the modules to teachers at the orientation sessions and also as a result of preparing the field test materials to send to the teachers. They were personally interested in seeing the data and in relating those data to their knowledge of the schools and the modules.

Obviously, the more knowledge assimilated with regard to a given module, the more confidently the revision to be performed on that module could be prescribed. It was decided to design the data processing and analysis procedures so as to maximize human contact with the data but to make those human operations as efficient as possible. Instead of designing machine-scorable forms, we scored tests by hand with machine assistance where appropriate.

After completing a folder on a given module, the field test person wrote a summary statement of his or her observations, impressions, and recommendations regarding revision of that module. The summary statement, in turn, served as additional information for the revision staff.

Scoring the module tests was the largest of the data processing tasks. The plan was for students to answer multiple-choice questions on the test itself rather than using a separate answer sheet. This was a deliberate choice to ensure that data were obtained whether or not students were familiar with the use of answer sheets. Teachers were given both the answer key and the key for linking items to objectives so that they could score the tests themselves for instructional purposes. Nevertheless, all tests were scored again by project field test staff. This scoring was done by typing item responses on an on-line remote terminal at AIR. The computer scored the test, computed summary statistics on the whole class and on male and female subgroups, and computed item analyses based upon the performance of the whole class. These results were printed out immediately at the terminal. Thus, the field test staff person saw each student's test while typing the item responses, a process that took about a minute per test. He or she also immediately saw the summary statistics and item analyses and therefore could easily associate them with the raw data.

This computer-assisted hand-scoring procedure, of course, took a little more time than would have been required to hand check and code machine-scorable answer sheets; however, it turned out that item responses could be processed in this manner without adding new staff. The process permitted staff to be involved with raw data as a source of information and impressions about the effectiveness of the module in a given class.

Analysis of covariance computations were also done via remote terminal at AIR after data were returned from relevant treatment and control classrooms. The variables of primary interest were treatment versus control main effects and treatment-by-sex interaction effects. The latter provided information relevant to possible sex biases in the module. Many assumptions about the data underlie analysis of covariance procedures, and further problems are associated

with unequal within-cell N's. Regression analysis techniques in analysis of covariance help in clarifying some of those assumptions and in sorting out those most pertinent to the design at hand. Nevertheless, because of the assumptions involved, the analysis of covariance results were not examined and interpreted in isolation from the many other field test data. Instead, all the data regarding a given module were examined together in order to cross-check emerging hypotheses about that module.

The total set of data on each given module was assembled in a folder for the revision staff. Included was information about the module's appeal to teachers at the teacher orientation workshops where that module was shown. This was indicated by the number of teachers who selected it as one of the three modules they would like to teach. Information about the field test districts for that module included region of the country, type of community, median family income, ethnic composition, and rated interest in career education. All teachers' ratings and comments on the Module Evaluation Form were recorded by the field test staff on a single form and color coded to teachers for easy comparison. Teachers' comments were rewritten verbatim on the same form and also color coded to teachers. A single form with all the teachers' ratings and verbatim comments was prepared for each learning activity in the module. The field test staff prepared another form showing the number and percent of students in each class who answered yes, maybe, and no to each of the six questions on the Student Evaluation Form. Every comment was rewritten verbatim and coded to indicate the class. These were attached to the Student Evaluation summary sheet. Number, mean, and standard deviation on pretest and posttest were shown separately for each treatment and control class. Also shown were the percent of students in each class who selected each alternative on each item of the pretest and posttest. Finally, the Teacher Guides with teachers' comments were included in the folder.

In retrospect, it must now be asked whether the field test data were effectively organized, assimilated, understood, and used. Should it be done differently next time? Could these or similar procedures be recommended to others engaged in curriculum evaluation?

The response to these questions is that the process of living with the data through various stages of hand processing is extremely valuable under certain conditions and with regard to some goals. The process would not be possible under other conditions and would not be of any value with regard to other goals. Personal involvement is pertinent if the goal is to revise and polish the final curriculum product. If the field test is simply to check the effectiveness of the curriculum and no revision is to be done, then any required hand processing is only a clerical chore and should be minimized.

It was necessary that the field test staff know the curriculum materials well and know the school settings and teachers who field tested those materials. Otherwise, the hours of data processing would have been without context and would have been meaningless. Members of the field test staff were unusually competent, intelligent, energetic, and persistent; these were truly necessary conditions to the success of the process. Without these traits, the tasks involved in this large field test effort would have been overwhelming.

MODULE EVALUATION

Module: _____ Date Evaluated: _____ By: _____

A. Are the materials appropriate?

relevance to career education:	() relevant	()	() questionable	()	() irrelevant
consonance of activities and objectives:	() fit together	()	() some fit	()	() not related
reading/vocabulary level for target grade range:	() appropriate	()	() somewhat high	()	() much too high
grade level:	() module appropriate for this grade		more appropriate for _____		

COMMENTS

B. Are the activities worthwhile?

motivation:	() highly motivating	()	() satisfactory, but uninspiring	()	() distinctly non-motivating
flexibility of adaptation to different students:	() flexible	()	() questionable	()	() inflexible
sequence of activities:	() very orderly	()	() fairly orderly	()	() haphazard

COMMENTS (including suggestions for additional activities/options)

C. What is the quality of the materials?

layout and pictorial matter:	() stimulating	()	() satisfactory	()	() uninviting
accuracy of information:	() accurate	()	() minor misrepresentation	()	() major misrepresentation and errors
bias:	() equally attractive to all subgroups	()	() some bias	()	() slanted toward or against one sex or group
comprehensiveness:	() adequate coverage of topic	()	() neglect of some aspects	()	() large inadequacies
clarity:	() clear	()	() ambiguous	()	() confusing
evaluation materials:	() reflect module content	()	() somewhat reflect content	()	() fail to reflect content

COMMENTS

D. How feasible is implementation?

infusion into the curriculum:	() fits fairly well in present form	()	() fits with minor changes	()	() not usable in present form
primarily infusible into:	() language arts	() mathematics	() science	() social studies	() other; specify _____
teacher preparation time:	() minimal	()	() reasonable	()	() excessive
special training for teachers:	() none required	()	() some training needed	()	() requires considerable training
management of classroom activities:	() easy	()	() moderately hard	()	() difficult

COMMENTS

E. What is the overall acceptability of this module?

to teachers:	() highly acceptable	()	() acceptable	()	() unacceptable
to students:	() highly acceptable	()	() acceptable	()	() unacceptable

COMMENTS

F. Would you use this module again?

() yes, with no revisions	()	() yes, with a few changes	()	() no
-------------------------------	-----	--------------------------------	-----	-----------

COMMENTS

ACTIVITY EVALUATION

Module: _____ Activity: _____ Date: _____

☐ I used this activity.

☐ I used only part of this activity;

☐ or none at all;

My preparation time was:

☐ negligible ☐ 15 min. ☐ 30 min. ☐ 45 min. ☐ hour or more

Total student class time for this activity was:

☐ 30 minutes or less ☐ 31 - 45 minutes ☐ 46 - 60 minutes ☐ between 1-2 hours ☐ more than 2 hours

If applicable, the student workbook/activity sheets were:

☐ very useful ☐ useful ☐ useless

The student's reaction to the material was one of:

☐ considerable interest ☐ some interest ☐ boredom, anathy

The overall acceptability of the activity was:

☐ highly acceptable ☐ acceptable ☐ unacceptable

due to:

- ☐ excessive length,
- ☐ probable student boredom,
- ☐ inappropriate, extraneous content or activities,
- ☐ difficulty of teaching it,
- ☐ material already taught elsewhere,
- ☐ other; please specify _____

I modified this activity by:

- ☐ lengthening it,
- ☐ shortening it,
- ☐ substituting another activity,
- ☐ changing the sequence or steps,
- ☐ substituting materials.

Appendix B

COMMENTS:

PROJECT STAFF

Project Management:

James Dunn, Project Director
Dewey Lipe, Associate Project Director
Gordon McLeod, Director of Revision
Jacqueline Haveman, Director of Field Testing
Richard Johnson, Director of Instrumentation
and Data Analysis

Revision Staff:

Carol Bain
Ann Bouie
Barbara Dodson
Carol Kaplan
Deborah Plumley
Irene Preiss
Jeri Seese-Hayes
Jeanette Wheeler
Mary Willis

Field Test and Data Analysis Staff:

Karen Bianchini
Vicki Chao
Janie Curry
Dale Perkins
Debra Thomason

Administration and Production Staff:

Caroline Dillman
Jay Egan
Jane Spangler
Terrie Willis
Irene Yurash

37-A